

ENGLISH: BEST FOR(E) PLAY WITH WORDS

A. ROSS ECKLER

Morristown, New Jersey

The following talk was presented at the international conference celebrating the 20th anniversary of the American Society for Geolinguistics at New York University on April 20 1985

Let me put my cards on the table at the start of this talk. I'm a language chauvinist—I believe English is the greatest of all languages when it comes to wordplay. Of course, sometimes I wake up in the middle of the night with the uncomfortable thought that I am a product of my environment, and that if I were a Frenchman I'd be saying the same thing about French. But I really do feel that a case can be made for English on several grounds—some well-established, others needing more verification. To summarize:

- The number of English words
- The polyglot nature of English words
- The statistical structure of English words
- The syntax of English

I take a broad view of what constitutes wordplay—from letter-manipulation to form palindromes and word squares, or through puns and cryptic crossword puzzle clues. Words are collections of letters, collections of sounds, or carriers of meaning, and wordplay can occur at all three levels.

The number of words in English is quite ill-defined. There are perhaps 400,000 words and phrases in the second edition of Merriam Webster's Unabridged, not counting plurals of nouns, past tenses of verbs, and participles. In addition, there are many words in technical and special dictionaries. More to the point, there are lots of words in English-language test that don't appear in any dictionary—neologisms, hyphenated coinages, inventions with prefixes or suffixes (such as *-wise* words or *-gate* words). If you doubt this, look at Kučera and Francis' million-word sample of American English (as of 1962) which lists hundreds of such words among its 35,000 different types. I suspect that the true count for English is several million—and you should add a million more if you count surnames of people living in the United States as "words". I know of no other language that lays claim to such a stock of words.

The size of the vocabulary directly affects wordplay in at least two ways. First, there is a greater likelihood that one word can be transmuted into another by a letter-change, a letter-omission, a letter-insertion, or a letter-pair inversion (such as MINUTE-MINUET). It is interesting to note that the Reader's Digest has for many years had a "Pardon, Your Slip Is Showing" column reporting on amusing typos of this nature, but no analogous column in its foreign-language edition.

The second way size affects wordplay involves a mathematical argument. If you have (say) 50,000 eleven-letter words instead of 24,000, the chance of finding a word which can be transposed into another (like ENUMERATION into MOUNTAINEER) is quadrupled, not doubled. This multiplication effect becomes even more important in wordplay involving the cooperation of many words in a group—for example, in constructing a 9-by-9 word square, doubling the pool of words raises the number of possible squares by 500! In English, members of

the National Puzzlers' League have constructed over 1000 9-by-9 word squares in the last century (many using very obscure words), but I know of only a handful in French or any other language. I hasten to add that I don't think English is that much better; no doubt diligence in searching also plays a part.

Another way to look at the size of the English stockpile is through synonymy. One hears of statements such as "there are forty different Finnish words for various snow conditions" which English doesn't have, but the all-time champion for synonymy appears to be the English word "inebriated" or "drunk". Several people have collected sets of words or phrases for this, including such worthies as Benjamin Franklin, the critic Edmund Wilson, and the writer Henry L. Mencken, but it remained for Paul Dickson to compile the definitive list of 2,231 synonyms from "abuzz" to "zozzled". (These are published in his marvelous book entitled *Words*, published by Delacorte Press in 1982.) If any other language can top this, I'd love to hear about it.

Large as the English word-stock is, it is not perfect. There are a number of situations which other languages have a single word for, but which English must express by a circumlocution. Some of these were described by Dmitri Borgmann in the February 1970 *Word Ways*. For example, the English phrase ONE AND ONE HALF is succinctly described by Polish POLTORA, German ANDERTHALB or Latin SESQUIALTER. (In fact the Polish even have the word POLOSMA for SEVEN AND ONE HALF!) Similarly, English THE DAY BEFORE YESTERDAY is more compactly expressed by German VORGESTERN or Spanish ANTEAYER, and English THE DAY AFTER TOMORROW, by German UBERMORGEN, Italian DOPODOMANI or Polish POJUTRZE. To express relationships, German uses GESCHWISTER for BROTHERS AND SISTERS, Polish uses STRYJ for the paternal uncle and WUJ for the maternal uncle, and SZWAGROSTWO for one's husband's brother and his wife.

Other interesting words are Dutch KWELDER (the land on the outside of a dike), Provençal UBAC (the sunless north side of a mountain), Turkish HARFENDAZ (someone who makes insulting remarks to women in the street), Brazilian Portuguese MATAO (a jockey who crowds the others against a fence) and the Russian RAZNOGLAZY (having eyes of different colors).

But in defense I offer some superbly specialized dictionary-sanctioned English words:

UCALEGON a neighbor whose house is on fire
SEREIN a fine rain falling from an apparently cloudless sky near sunset
QUALTAGH the first person you meet on leaving home on a particular day (at the New Year)
QUASQUICENTENNIAL 125th anniversary
SHREWSTRUCK struck by a shrew (the animal, not a female scold)
NOSARIAN one who argues that there is no limit to the possible largeness of a nose

My point is this—English may lack words for some concepts, but it is at least as inventive as other languages in coining words for specialized needs.

The second reason for English's supremacy is wordplay is its polyglot nature—not only is it a fusion of Germanic and Romance languages, but an inveterate borrower of words from others languages as needed. For wordplay, this furnishes a superb stock of weird specimens, enabling one to carry out such projects as a type collection of bigrams—all two-letter combinations from AA bazAAR to ZZ fuZZy. Many of the words with Q-not-followed-by-a-U are direct borrowings from the Arabic: ZAQQUM (a bitter fruit tree), WAQF (a charitable trust), QOBAR (a Nile dry fog), QAID (a local North African official), FAQIH (a Muslim theologian), FIQH (Muslim

jurisprudence based on theology), TALUQDARI (a landholding tenure in India), and TAQLID (uncritical acceptance of Muslim theology).

Though I can't prove it, I suspect that the polyglot nature of English is also the reason for a fertile source of wordplay—homonyms, words spelled differently but pronounced the same (as “bear” and “bear”), and heteronyms, words pronounced differently but spelled the same (as “polish” and “Polish”). If sound and spelling are consistent, as they are to a far greater extent in most other languages, Homonyms and heteronyms are correspondingly reduced, and you don't get such delightful puns as the story about the mother whose cowboy offspring named their ranch Focus—where the sons raise meat.

I come now to the third reason why I think English is a superior wordplay language—the statistical structure of English words. This is a much less-known field than language size or diversity, and much work needs to be done to clarify differences among languages and their relevance to wordplay.

As a retired statistician, I have a somewhat different view of Language than most of my audience. At least some of you have heard Eddington's claim, first made in 1927, that if you had an army of monkeys strumming on typewriters you could eventually reproduce all the books in the British Museum. What he didn't say was the time it would take—for example, it would take a monkey typing ten characters per second about 100 decillion years to produce the first nine words of Hamlet's “to be or not to be” soliloquy. Long before the British Museum had been duplicated, the entire universe would be filled with books full of gibberish, and there would be no way of sorting the wheat from the chaff.

However, the concept *is* useful—can one mimic a language by its letter frequencies? Better yet, what about using its conditional frequencies (given that an E has appeared, what is the probability of each possible alphabetic letter following it in a randomly-selected text?). There are two interesting articles on this topic, one by William R. Bennett jr in the Nov/Dec 1977 journal *The American Scientist*, and one by Brian Hayes in the November 1983 *Scientific American* magazine. I can't go into details here, but they discovered that languages are sufficiently different in their letter-frequency statistics that it is not difficult for a person to tell the difference between second-order simulation of Latin, English, German, French, etc. If one goes to fourth-order statistics (each letter determined by its relative frequency of appearance following *three* specified earlier letters), one can actually differentiate among individual authors such as Shakespeare, Poe, Hemingway or Faulkner, based on their statistics. Of course, it is hardly surprising that language differences show up with less conditionality than author differences—but what surprises me is how little conditionality is needed to mimic the “feel” of a language or an author.

What letter-frequency statistics enhance or impede the suitability of a language for wordplay? The answer is not clear. It is known that it is harder to construct Italian crossword puzzles than English ones. The letter distribution of the final letters in Italian words—Italian has a higher percentage of vowels than English does—is said to create difficulties in crossword construction because one needs words with many vowels to match a series of word-ends. Is this why there are no 9-by-9 Italian word squares?

On the other hand, Hawaiian has only 12 different letters with five vowels, which leads to a far larger number of long all-vowel words in Hawaiian than in English. It also generates such curiosities as long cadences—the same letter repeated at even intervals like I in DIVISIBILITY. English has appropriated a notable Hawaiian example, HUMUHUMUNUKUNUKUAPUAA (the Hawaiian triggerfish).

Like German, Finnish is an agglutinative language, with long words consisting of many components. This may be the reason why its longest palindrome, SAIPPUAKAUPPIAS, exceeds the longest English palindrome, KINNIKINNIK, by four letters.

The final reason why English is superior for wordplay is its syntax. One aspect of this is exploited in cryptic crosswords—a word can be several different parts of speech, leading to ambiguous statements like FLYING PLANES CAN BE DANGEROUS, where “flying” is either a gerund or an adjective. Petr Beckmann, in his book *The Structure of Language—A New Approach* (Golem Press, 1972), views natural languages as examples of error-detecting and error-correcting codes—that is, syntax provides various words and word-endings that do not carry information but simply are used to provide a check for consistency, like the verb ending S in the third person singular, or gender pronouns. Beckmann argues that English has a much flimsier system of such consistency-checks than other languages (almost no case endings compared to Latin or Russian, no gender for nouns, few verb-endings corresponding to the subject, etc.). Instead, English resolves ambiguity by much stricter rules of word order.

Beckmann shows with numerous examples how English syntax gets into trouble with ambiguity when only a few words are omitted, as commonly occurs with headlines in newspapers:

FRENCH ARMYH COOKJS (are) WOMEN
MAYS PLAYS WITH (in spite of? using?) INJURED FINGER
MAN REFUSES TO GIVE UP (his) BITING DOG

As dessert, I describe a study by Arnold Rosenberg of IBM, published in *Linguisticae Investigationes* in 1979. He attempts to answer the question of what is the hardest language to learn. Language A is harder than Language B if (1) in Language B, the assertion “It is A to me” exists, meaning that “A is unintelligible” or (2) Language A is harder than Language C, and Language C in turn is harder than B (transitivity).

For example, in English, one says “It’s Greek to me”, so Greek is rated harder than English. In turn, the Greeks have an expression “It appears Chinese to me”, so by Rule 2 Chinese is harder than English. In fact, his studies show that Chinese is the hands-down winner, being rated harder by Spanish, Greek, Polish, Hebrew, Finnish, Estonian, Flemish, Hungarian, Romanian, Russian, Serbo-Croatian, Swiss German and Filipino! In addition it indirectly dominates Czech, German, English, Dutch, Afrikaans, Portuguese and French. So to whom do the Chinese turn? To Heaven, as evinced by their saying “It’s heavenly script to me”. The Danes picked Volapük as their dominant language—an artificial language of the 1880s supposed to be easy, but eclipsed by Esperanto.

It is interesting that there is a closed cycle in these language chains:

Turkish: if I understood that, I’d be an Arab
Arabic: it’s Persian to me
Persian: did you say it in Turkish?

It’s perhaps significant that the Tower of Babel was built in this part of the world. It’s also time for me to quit before I chew on my own tail, like the Ouroboros Worm featured in Eddison’s 1922 science-fiction story!